Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 12, No. 2, 2024

Characterization of Escherichia coli as a Recombinant Protein Expression Host and its Optimization

Kang Yu, Congjing Cao, Luwen Huang, Enxu Wang *

Bloomage Biotech (Tianjin) Technology Co., LTD, Tianjin 300000, China

*** Corresponding author:** Enxu Wang (Email: wangenxu@bloomagebiotech.com)

Abstract: Engineered *Escherichia coli* serves as an efficient platform for the production of a diverse array of recombinant proteins, thereby addressing the pharmaceutical industry's demand for high-quality biological agents. Nevertheless, challenges related to protein folding and the complexities involved in product purification must be acknowledged. This review delineates the advantages and disadvantages of utilizing *E. coli* as an expression host and further investigates various strategies aimed at optimizing and enhancing the *E. coli* expression system to elevate the yield, purity, and biological activity of recombinant proteins, ultimately supporting advancements in biopharmaceuticals and other relevant fields.

Keywords: *Escherichia Coli*; Recombinant Protein; Expression Host; Optimization Strategy.

1. Introduction

With the rapid development of molecular biology and proteomics, the genetic manipulation technology of exogenous gene expression is becoming more and more mature. Expression system is the core of exogenous gene expression, and commonly used expression systems are generally model organisms, including eukaryotic and prokaryotic expression systems, among which eukaryotic systems include mammalian cell expression system, plant expression system, insect baculovirus expression vector system and yeast expression system, and prokaryotic expression system is mainly *E. coli* expression system. *E. coli* is one of the most widely used prokaryotic expression system and the earliest exogenous gene expression system to be studied, with a clear genetic background, fast growth, easy to achieve high density cultivation, low cost, high yield and incomparable superiority compared with other expression systems, it is the most widely used expression system in commercial production, and it has achieved great scientific research value and economic benefits [1].

The *E. coli* expression system is now widely used for the expression production of a variety of protein/peptide drugs and biochemicals, including: recombinant human insulin, a2b-type interferon, ranibizumab, purple bacillin and peony skin glucoside[2]. According to statistics, 26% of recombinant protein drugs approved for marketing by the U.S. FDA and the European EMA from 1986-2018 came from *E. coli*. Meanwhile, genetically engineered vaccines expressed through *E. coli* are now entering the market or in the clinical trial stage, such as hepatitis E vaccine, human papillomavirus vaccine, influenza A vaccine and so on. Common *E. coli* expression systems include BL21 series, JM109 series, W3110 series and K802 series, etc. Among them, *E. coli* BL21 (DE3) strain is one of the most widely used strains in recombinant protein expression research, BL21 (DE3) is obtained from the derivatives of *E. coli* B series and K-12 series through genetic mutations such as P1 transduction. These strains are usually host protease-deficient to ensure that the exogenous proteins are not degraded during the expression process and to maintain the stability of expression [1,2].

2. Advantage of Escherichia Coli as an Expression Host

As an expression host, *Escherichia coli* has been widely used in the field of biopharmaceutical because of its many advantages.

The study of *Escherichia coli* has a long history, with a lot of data and experience accumulation. At the same time, its genetic background is quite clear, and researchers have a very in-depth understanding of the *E. coli* genome. This allows scientists to easily perform gene cloning, expression regulation, and genetic engineering operations. By using multiple vectors and expression systems, *E. coli* can effectively express foreign genes and produce target proteins [3,4]. Due to the widespread use of *E. coli* in recombinant protein production, a large number of molecular biology tools and techniques have been established. Researchers can utilize a wealth of expression plasmids, promoters, screening markers, and culture conditions to optimize the production of target proteins. The availability of these technologies allows researchers to more effectively address the various challenges encountered during protein expression [5,6].

E. coli is good at producing non-glycosylated proteins, which is particularly important for many biotechnology applications. Given the complexity of mammalian cell systems in glycation modification, many researchers have chosen to use *E. coli* to produce recombinant proteins that do not require glycation, thereby avoiding the associated complexity and cost [7,8]. *Escherichia coli* has a very fast growth rate, and this high growth rate enables *Escherichia coli* to reach a high density of cells in a short time, thereby rapidly producing a large number of recombinant proteins [9]. The medium required for *E. coli* is inexpensive and easy to prepare, and the culture process is often simpler and more convenient than other systems [7,10]. *E. coli* can adapt to a variety of growth conditions, including different temperatures, pH, and oxygen concentrations. This adaptability enables efficient production of target proteins under different production environments. In addition, by means of genetic engineering, its metabolic pathway can be further optimized to enhance the yield and quality of recombinant proteins [6,9].

These properties make it an important tool in the field of biotechnology and also make the use of *E. coli* in industrial production a cost-effective option.

3. Challenges of Escherichia Coli as an Expression Host

Despite *E. coli*'s strengths in genetic engineering and protein expression, it still faces many challenges in the folding and aggregation of recombinant proteins.

When *E. coli* expresses recombinant protein at high levels, it is easy to cause the protein to aggregate into inclusion bodies. This phenomenon is usually caused by factors such as high temperature, high induced concentration and strong start-up subsystem [11,12]. Under these conditions, the translation rate of the protein increases, resulting in insufficient folding time, resulting in the formation of non-functional aggregates. Some recombinant proteins have complex structures or are highly toxic, and their intracellular expression may burden host cells and increase the risk of misfolding [6]. In addition, many heteropeptides, when expressed in *E. coli*, fail to fold into their natural state and are subsequently degraded by cells or aggregated into insoluble forms [13,14]. Changes in the intracellular environment, such as pH, salt concentration and temperature, also affect the folding process of proteins [15,16].

4. Strategies for Optimization

4.1. Optimization of Clone Design

Commonly used *E. coli* carriers are pET, pCold, pBV and pGEX series. The component composition and promoter type of the vector affect the gene expression level and solubility. It is very important to select suitable expression vector to improve protein expression. Differences in codon use between *E. coli* and eukaryotes may result in inadequate expression of foreign genes. Therefore, codon optimization is an effective method to improve the expression of recombinant protein. By adjusting the codon of the target gene to better match the preference codon of *Escherichia coli*, the expression level of the protein can be significantly improved [17]. It has been found that replacing the rare arginine codon with the more commonly used codon increases the production of human interferon 2b in *E. coli* by about 10 times [18]. Fusion tags are often used to improve protein solubility and purification processes. In clonal design, some tags are often fused to the N-terminal or C-terminal of the target protein, such as His, GST, MBP, NusA, Trx, and SUMO. However, the main challenge in protein production is how to split and remove fusion tags during the different steps of purification. To achieve this, multiple cleavage sites have been designed for proper protein expression and cleavage. In addition, in some cases, the target protein may result in low expression due to misfolding or aggregation, in which case, co-expression of molecular chaperones (such as DnaK, GroEL, GroES, etc.) can be considered to promote the correct folding of the target protein, improve solubility and reduce the formation of inclusion bodies. This strategy has been shown to increase the production and solubility of difficult-to-express proteins [6,19,20].

4.2. Optimization of Culture and Induction

Optimizing induction conditions in *Escherichia coli* is a complex and important process involving the comprehensive consideration of several factors. The expression efficiency of recombinant protein could be significantly improved by selecting the inducer and its concentration, induction time, culture temperature and medium composition.

The type and concentration of the inducer have significant influence on protein expression, and it is very important to select the appropriate inducer and its concentration to obtain high yield of recombinant protein. The most commonly used inducer is isopropyl-beta-d-thiogalactoside (IPTG). In general, excessively high concentrations of IPTG may lead to the accumulation of protein folding precursors and inclusion body formation. The use of lower concentrations of inducers can reduce the cell burden, thereby increasing the solubility and expression of the target protein. However, too low concentration may affect protein production. Therefore, it is necessary to determine the optimal IPTG concentration by experiment. In addition, in addition to IPTG, other inducers such as lactose, tryptophan, and arabinose can also be considered. Lactose autoinduction has mild characteristics and is suitable for the efficient expression of certain proteins. For example, studies have shown that when 10% lactose is used as an inducer, it is suitable for induction after the culture reaches a specific optical density (OD), which can effectively improve the protein expression [21].

The choice of induction time is also very important. In general, induction should be carried out during the logarithmic phase of bacterial growth, when the cell activity is higher and is better able to respond to induction. For example, induction at OD values of 1.0 to 1.5 can significantly improve the expression efficiency of recombinant proteins [22]. Too early or too late induction can result in insufficient protein expression or the formation of aggregates, which can affect the quality of the final product.

Certain chemical additives, such as glycerin, trehalose, etc., can improve intracellular viscosity and help improve the folding state of recombinant proteins. These substances increase the production of soluble proteins by altering the microenvironment within the cell and reducing the formation of aggregates [16,23].

Culture temperature is another important factor affecting protein expression. Induction at lower temperatures generally improves the solubility of recombinant proteins and reduces the formation of inclusion bodies. Lowering the induction temperature contributes to the metabolic burden of the cell, which in turn promotes the correct folding and stability of the target protein [19,25]. Certain proteins tend to aggregate and dissolubilize at 37°C, and setting the induction temperature at 30°C can achieve better expression results in some cases, while reducing the temperature to 18°C may also show better results in other cases.

The growth rate and expression level of recombinant protein were affected by different medium components. For example, the addition of certain amino acids or vitamins can promote the growth and metabolism of cells, thus improving the expression effect of proteins. In addition, optimizing the ratio of carbon and nitrogen sources in the medium is also an important strategy to improve protein expression [25].

4.3. Pilot Test Optimization

It must be mentioned that a major challenge that laboratory research results must face is scale-up production, which usually needs to follow a certain scale-up strategy through pilot tests, optimize fermentation conditions according to fermentation processes and equipment, and develop

separation, purification, drying and other processes in line with large-scale production to ensure that the stability and consistency of the final product is maintained in reactors of different sizes.

5. Conclusion

Escherichia coli has been widely used in the field of recombinant protein, and as one of the most commonly used expression systems in genetic engineering, it has shown remarkable advantages and potential. The application of *E. coli* in the field of recombinant proteins will continue to expand and optimize. With the introduction of high-throughput screening technology, researchers can more systematically explore and determine the optimal expression conditions to further improve protein production and activity. At the same time, the limitations of *E. coli* expression system can be improved through genetic engineering, expression vector optimization and culture conditions adjustment. With the continuous progress and innovation of technology, it will play a more important role in the development of the biomedical industry.

Author Contribution Statement

All authors listed have significantly contributed to the development and the writing of this article.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by Bloomage Biotech (Tianjin) Technology Co., LTD, project 24ZYCGSY00560.

References

[1] Hayat SMG, Farahani N, Golichenari B, Sahebkar A. Recombinant Protein Expression in Escherichia coli (E. coli): What We Need to Know. Curr Pharm Des. 2018;24(6):718-725.

[2] Ahmad I, Nawaz N, Darwesh NM, Ur Rahman S, Mustafa MZ, Khan SB, Patching SG. Overcoming challenges for amplified expression of recombinant proteins using Escherichia coli. Protein Expr Purif. 2018 Apr; 144:12-18.

[3] Gopal GJ, Kumar A. Strategies for the production of recombinant protein in Escherichia coli. Protein J. 2013 Aug; 32 (6): 419-425.

[4] Saroha P, Patil RS, Rathore AS. Recent advancements in soluble expression of recombinant antibody fragments in microbial host systems. Prep Biochem Biotechnol. 2024 Aug 28:1-10.

[5] Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. Front Microbiol. 2014 Apr 17; 5:172.

[6] Zhang ZX, Nong FT, Wang YZ, Yan CX, Gu Y, Song P, Sun XM. Strategies for efficient production of recombinant proteins in Escherichia coli: alleviating the host burden and enhancing protein activity. Microb Cell Fact. 2022 Sep 15;21(1):191.

[7] Huang CJ, Lin H, Yang X. Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. J Ind Microbiol Biotechnol. 2012 Mar;39 (3): 383-399.

[8] Gupta V, Sengupta M, Prakash J, Tripathy BC. Production of Recombinant Pharmaceutical Proteins. Basic and Applied Aspects of Biotechnology. 2016 Oct 23:77–101.

[9] Rosano GL, Morales ES, Ceccarelli EA. New tools for recombinant protein production in Escherichia coli: A 5-year update. Protein Sci. 2019 Aug;28(8):1412-1422.

[10] İncir İ, Kaplan Ö. Escherichia coli as a versatile cell factory: Advances and challenges in recombinant protein production. Protein Expr Purif. 2024 Jul; 219:106463.

[11] Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK. Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. Microb Cell Fact. 2015 Mar 25; 14:41.

[12] Baneyx F. Recombinant protein expression in Escherichia coli. Curr Opin Biotechnol. 1999 Oct;10(5):411-421.

[13] Georgiou G, Valax P. Expression of correctly folded proteins in Escherichia coli. Curr Opin Biotechnol. 1996 Apr;7(2):190-197.

[14] Falak S, Sajed M, Rashid N. Strategies to enhance soluble production of heterologous proteins in Escherichia coli. Biologia (2022) 77:893–905.

[15] Schramm FD, Schroeder K, Jonas K. Protein aggregation in bacteria. FEMS Microbiol Rev. 2020 Jan 1;44(1):54-72.

[16] de Marco A, Vigh L, Diamant S, Goloubinoff P. Native folding of aggregation-prone recombinant proteins in Escherichia coli by osmolytes, plasmid- or benzyl alcohol-overexpressed molecular chaperones. Cell Stress Chaperones. 2005 Winter;10 (4): 329-339.

[17] Maldonado LM, Hernández VE, Rivero EM, Barba de la Rosa AP, Flores JL, Acevedo LG, De León Rodríguez A. Optimization of culture conditions for a synthetic gene expression in Escherichia coli using response surface methodology: the case of human interferon beta. Biomol Eng. 2007 Jun;24(2):217-222.

[18] Valente CA, Prazeres DM, Cabral JM, Monteiro GA. Translational features of human alpha 2b interferon production in Escherichia coli. Appl Environ Microbiol. 2004 Aug;70 (8): 5033-5036.

[19] Kaur J, Kumar A, Kaur J. Strategies for optimization of heterologous protein expression in E. coli: Roadblocks and reinforcements. Int J Biol Macromol. 2018 Jan; 106:803-822.

[20] Ge J, Wang X, Bai Y, Wang Y, Wang Y, Tu T, Qin X, Su X, Luo H, Yao B, Huang H, Zhang J. Engineering Escherichia coli for efficient assembly of heme proteins. Microb Cell Fact. 2023 Mar 28;22(1):59.

[21] Mühlmann M, Forsten E, Noack S, Büchs J. Optimizing recombinant protein expression via automated induction profiling in microtiter plates at different temperatures. Microb Cell Fact. 2017 Nov 28;16(1):220.

[22] Rigi G, Rostami A, Ghomi H, Ahmadian G, Mirbagheri VS, Jeiranikhameneh M, Vahed M, Rahimi S. Optimization of expression, purification and secretion of functional recombinant human growth hormone in Escherichia coli using modified staphylococcal protein a signal peptide. BMC Biotechnol. 2021 Aug 16;21(1):51.

[23] Vincentelli R, Romier C. Expression in Escherichia coli: becoming faster and more complex. Curr Opin Struct Biol. 2013 Jun;23(3):326-334.

[24] Bahreini E, Aghaiypour K, Abbasalipourkabir R, Goodarzi MT, Saidijam M, Safavieh SS. An optimized protocol for overproduction of recombinant protein expression in Escherichia coli. Prep Biochem Biotechnol. 2014;44(5):510-528.

[25] Zhou Y, Lu Z, Wang X, Selvaraj JN, Zhang G. Genetic engineering modification and fermentation optimization for extracellular production of recombinant proteins using Escherichia coli. Appl Microbiol Biotechnol. 2018 Feb;102(4): 1545-1556.